Channel correction via quantum erasure

Francesco Buscemi
ERATO-SORST Quantum Computation and Information Project, Japan Science and Technology Agency, Daini Hongo White Building 201, 5-28-3 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.[*]

(Dated: February 1, 2008)

By exploiting a generalization of recent results on environment-assisted channel correction, we show that, whenever a quantum system undergoes a channel realized as an interaction with a probe, the more efficiently the information about the input state can be erased from the probe, the higher is the corresponding entanglement fidelity of the corrected channel, and vice-versa. The present analysis applies also to channels for which perfect quantum erasure is impossible, thus extending the original quantum eraser arrangement, and naturally embodies a general information-disturbance tradeoff.

In quantum cryptography, the ability of faithfully transmit arbitrary quantum states between two parties is a sufficient condition to achieve privacy with respect to a third malicious party, which is usually described as the purifying environment. The converse statement is not true, in the sense that it has been recently proved that it is possible to distill private states from channels which have zero capacity [1]. There are common situations, however, in which the third party is not malicious, but (in some degree) helpful. This is the case of *environment-assisted channel correction* [2]. This Letter shows that, in this case, privacy and quantum capacity are equivalent quantities: the assisting party can choose to give up (to *erase*) some of her information about the transmitted states in order to help the other two parties to enhance the capacity of their quantum communication, and, the more she erases, the better is the corresponding assisted quantum capacity.

A paramount *ante litteram* example of assisted channel correction is the *quantum eraser* [3], that is, a variety of the usual double-slit interference experiment with a single particle, in which it is possible to mark either particle- or wave-like (exploring also halfway [4]) properties of the beam by measuring one observable among a set of non-commuting observables of the probe, which, previous to the measurement, has been made suitably interact with the beam in order to store the *which-path* information. Such which-path information sits in the correlations established during the interaction between the particle and the probe. In particular, Ref. [5] extends such kind of duality relations to more general situations, in which a two-level system interacts with a probe in such a way that measurements on the latter result in a sorting of the former into sub-ensembles exhibiting particle- or wave-like characteristics. It holds that, the more which-path information the measurement collects from the probe, the less visible are the fringes in the conditional sub-ensembles, and vice-versa.

Another analogous situation can be recognized in the *partial teleportation* of an unknown quantum state. In the ideal scenario [6], the two parties share a maximally entangled state, and the assisting party performs a Bell measurement jointly on the unknown state and on her branch of the shared entangled resource, telling the result to the assisted party. In this case, the assisting party is left with no information about the input state, while the assisted party can perfectly recover it with unit probability. The imperfect situation [7] happens when the shared state is not maximally entangled, or when the measurement is not a complete Bell measurement. In this case, the teleportation is noisy, in the sense that it happens either probabilistically or with a fidelity smaller than one. Correspondingly, the assisting party is left with *some* information about the state to be teleported [7]. Also in this case one would say, closely mirroring the quantum eraser situation, that, the less information the assisting party collects about the unknown state, the better is the quality of the teleported state at the assisted party's side, and vice-versa.

The two examples of quantum eraser and partial teleportation suggest that, whenever a quantum system interacts with an assisting environment (or probe), it may be possible to restore coherence lost under the effect of a noisy channel by "erasing" from the probe—that is, by performing on the probe a measurement whose outcomes are as much independent as possible of—the information carried by the input system itself. In the present Letter we will show that it is indeed possible to formalize such an insight and to extend the mentioned approach on a general basis. More explicitly, we will consider general quantum evolutions, mathematically described as channels—i. e. completely positive trace-preserving maps [8]—acting on an input system and physically modeled as unitary interactions of the input system with a probe, the latter playing the role of a controllable environment. Then, by exploiting the theory of environment-assisted correction [2], we will derive two inequalities relating the amount of information (about the input state) extracted from the probe with the entanglement fidelity of the corresponding corrected channel, showing that perfect erasure—that is, null information—is equivalent to perfect correction and, even if perfect erasure is impossible, a robust tradeoff relation between information extraction and channel correction efficacy holds, and an optimal correction scheme can be explicitly constructed. In this sense we can think that a sort of *quantum erasure relation* holds valid in all conceivable situations, also providing, as a byproduct, a quite general information-disturbance tradeoff relation.

Environment-assisted channel correction.— Let us given a channel $\mathcal{E}$ acting on density matrices ρ defined on the (finite dimensional) input Hilbert space $\mathscr{H}_S$. Basically, there exist two equivalent ways to represent a channel, both highly non unique. The first one is the Kraus representation [8], that is,

$$\mathcal{E}(\rho) = \sum_k E_k \rho E_k^\dagger, \qquad (1)$$

where the operators $\{E_k\}_k$ satisfy normalization $\sum_k E_k^\dagger E_k = \mathbb{1}$. The second one, coming as a direct consequence of the Stinespring theorem [9], regards the channel as the average of an indirect measurement scheme, in which first the system interacts with an environment (described by the Hilbert space $\mathscr{H}_E$ and initialized in a pure state $|0\rangle_E$), and then a Positive-Operator–Valued Measure (POVM) $\mathbf{M} = \{M_j^E\}_j$, $M_j^E \geq 0\ \forall j$, $\sum_j M_j^E = \mathbb{1}$, is measured on the environment, in formula [10]

$$\mathcal{E}(\rho) = \sum_j \mathrm{Tr}_E[U(\rho \otimes |0\rangle\langle 0|_E)U^\dagger\ (\mathbb{1} \otimes M_j^E)]. \qquad (2)$$

(In order to keep the notation simple, we consider, without loss of generality, the output systems S and E to be equal to the input ones.) The task is to correct the channel $\mathcal{E}$. Without having access to the environment (and hence to the assisting indices j), the best one can do is described in Ref. [11], in which it is shown that, if the coherent information is close to the input entropy, it is possible to devise a correcting channel $\mathcal{C}^\rho$, depending in general also on the input state ρ, such that $\mathcal{C}^\rho \circ \mathcal{E}$ is close to the identity channel on the support of ρ.

The assisted scenario, first introduced in Ref. [2], is much more powerful, since we now allow the environment to be somehow "controllable" or "assisting", much like a probe, in the sense that we can control the measurement operators M_j^E and have access to the measurement outcomes j. This is the case, for example, of the previously mentioned quantum eraser and partial teleportation. The corrected channel then can take the form $\sum_j \mathcal{C}_j^\rho \circ \mathcal{E}_j$, where $\mathcal{E}_j(\rho) = \mathrm{Tr}_E[U(\rho \otimes |0\rangle\langle 0|_E)U^\dagger\ (\mathbb{1} \otimes M_j^E)]$. The probability of getting the j-th outcome is equal to $p(j) = \mathrm{Tr}[\mathcal{E}_j(\rho)]$, the conditional output state is $\sigma_j := \mathcal{E}_j(\rho)/p(j)$, and both depend explicitly on the input system state ρ and on the POVM $\mathbf{M} = \{M_j^E\}_j$. In practical situations the hypothesis of complete control on the environment is clearly too strong: in all these cases we can separate the environment system E into a controllable probe P and "the rest" R. Then the POVM operators M_j^E will have the form $M_j^E = M_j^P \otimes \mathbb{1}_R$. Notice however that the assisted scenario, as motivated in [2], when assuming for granted that the system interacts with a completely controllable environment only, mainly focuses on the *in principle* information dynamics involved in the overall process, and not on the practical feasibility of the correction scheme itself. Hence in the following, when we will need the "complete controllability hypothesis", we will explicitly call for it: as a rule, we will speak of "probe" if *complete* control is possible, while the word "environment" will be left for all other cases. Notice, moreover, that for a (finite) d-dimensional system $\mathscr{H}_S$, a d^2-dimensional probe suffices to realize whatever quantum evolution.

Information retrieval and erasure.— First of all, let us fix some notation. Given a channel $\mathcal{E}$ acting on states of the input system S, from Eq. (2) we can always construct the so-called *complementary* channel $\widetilde{\mathcal{E}}$, defined as $\widetilde{\mathcal{E}}(\rho) = \mathrm{Tr}_S[U(\rho \otimes |0\rangle\langle 0|_E)U^\dagger]$. It describes the output state of the environment given that the input state of the system was ρ. By Stinespring theorem [9], such a complementary channel is unique up to a partial isometry [12]. Hence, we can consider $\widetilde{\mathcal{E}}$ as being the canonical complementary channel. Moreover, given a channel $\mathcal{E}$ acting on states, there exists a unique *dual* channel $\mathcal{E}'$ acting on observables O, defined by the trace relation $\mathrm{Tr}[\mathcal{E}(\rho)\ O] = \mathrm{Tr}[\rho\ \mathcal{E}'(O)]$, for all ρ. The trace-preserving condition becomes a unit-preserving condition, i. e. $\mathcal{E}'(\mathbb{1}) = \mathbb{1}$. We then have four channels: the direct one, i. e. $\mathcal{E}$; the dual one, i. e. $\mathcal{E}'$; the complementary one, i. e. $\widetilde{\mathcal{E}}$; and the complementary dual one, i. e. $\widetilde{\mathcal{E}}'$. If we send through the channel $\mathcal{E}$ an ensemble of quantum states $\{\rho_i\}_i$, such that $\mathrm{Tr}[\rho_i] = p(i)$, $\sum_i \rho_i = \rho$, and $\mathrm{Tr}[\rho] = 1$, at the environment output branch will arrive $\{\widetilde{\mathcal{E}}(\rho_i)\}_i$. We then perform a measurement on them by using a POVM $\mathbf{M} = \{M_j^E\}_j$, thus obtaining a joint probability distribution $p(i,j) = \mathrm{Tr}[\widetilde{\mathcal{E}}(\rho_i)\ M_j^E] = \mathrm{Tr}[\rho_i\ \widetilde{\mathcal{E}}'(M_j^E)]$. In the following, we will consistently use the index i for the input ensemble and the index j for the environment outcomes. We now invoke the complete controllability hypothesis, and choose rank-one POVM elements, i. e. $M_j^E = |\phi_j\rangle\langle\phi_j|_E := \phi_j^E$. In fact, such a choice is necessary and sufficient to rule out the possibility of a *classical post-processing* of data [13], which could artificially reduce the information transmission. Since in the following we will be interested in the measurement *minimizing* the information transmission, the choice of a rank-one probe measurement is definitely the appropriate one. Correspondingly, the channel $\mathcal{E}$ gets decomposed into pure contractive maps $\mathcal{E}_j(\rho) = E_j \rho E_j^\dagger$, or in other words, the rank-one measurement $\mathbf{M}$ refines the channel $\mathcal{E}$ into a *pure instrument* [14].

In order to quantify the amount of information about the input ensemble $\{\rho_i\}$ that the measure of $\mathbf{M}$ retrieves from the probe, it is natural to compute the *mutual information* from $\vec{p}(i,j)$ as $I_{S:E}^{\mathbf{M}} = H(\vec{p}(i)) + H(\vec{p}(j)) - H(\vec{p}(i,j))$, where $H(\vec{q}(k)) = -\sum_k q(k) \log q(k)$ is the

Shannon entropy of the probability distribution $\vec{q}(k)$. If $I^{\mathbf{M}}_{S:E}$ is close to zero, then $\vec{p}(i,j) \approx \vec{p}(i)\vec{p}(j)$, that is, $p(i,j) \approx p(i)p(j)$ for all i,j, namely, the outcomes of the measurement $\mathbf{M}$ on the probe are almost independent of the input ensemble $\{\rho_i\}_i$. It means that the information transmission is poor, and if the same holds for all possible ensemble realizations $\{\rho_i\}_i$ of ρ, we say that the measurement $\mathbf{M}$ performs a good *erasure* with respect to the input state ρ. Here the mutual information equals the *relative entropy* $D(\vec{p}(i,j)\|\vec{p}(i)\vec{p}(j))$ between the joint distribution $\vec{p}(i,j)$ and the factorized one $\vec{p}(i)\vec{p}(j)$, where $D(\vec{r}(k)\|\vec{s}(k))$ is defined for two probability distributions $\vec{r}(k)$ and $\vec{s}(k)$ as $D(\vec{r}(k)\|\vec{s}(k)) = \sum_k r(k) \log r(k)/s(k)$, [15]. If $s(k) = 0$ for some k for which $r(k) > 0$, then $D(\vec{r}(k)\|\vec{s}(k))$ diverges. In our case, however, this will never be the case and the following inequalities will play a central role [15, 16, 17]

$$2^{-1}\|\vec{r}(k) - \vec{s}(k)\|_1^2 \le D(\vec{r}(k)\|\vec{s}(k)) \le \beta^{-1}\|\vec{r}(k) - \vec{s}(k)\|_1^2, \tag{3}$$

where $\beta = \min_k s(k) > 0$, and $\|\vec{r}(k) - \vec{s}(k)\|_1 = \sum_k |r(k) - s(k)|$.

Correction efficiency.— A useful quantity to judge the capability of a channel $\mathcal{E}$ in faithfully and coherently transmitting an input state ρ, is given by the *entanglement fidelity* [18] $F_e(\rho)$ defined as $F_e(\rho) = \mathrm{Tr}[|\Omega\rangle\langle\Omega|\,(\mathcal{E} \otimes \mathcal{I})(|\Omega\rangle\langle\Omega|)]$, where $\mathcal{I}$ is the identity (ideal) channel and $|\Omega\rangle$ is a purification of ρ. If $F_e(\rho)$ is close to one, then the channel $\mathcal{E}$ acts quite like the identity channel on the support of ρ, [18]. Starting from a Kraus decomposition as in Eq. (1), with few calculations we find that $F_e(\rho) = \sum_k |\mathrm{Tr}\,\rho E_k|^2$, and since it does not depend on the particular decomposition $\{E_k\}$ chosen, it is an *intrinsic* property of the channel. The following simple upper bound then comes from an application of a Cauchy-Schwartz–type inequality

$$F_e(\rho) \le \sum_k (\mathrm{Tr}\,|\rho E_k|)^2 := F_{e,\mathrm{a}}(\rho) \le 1, \tag{4}$$

where $|\rho E_k|$ is the positive part of the polar decomposition $\rho E_k = U_k^\rho |\rho E_k|$, for unitary U_k^ρ. With an environment-assisted correction scheme, it is indeed possible to reach such an upper bound, that we therefore call $F_{e,\mathrm{a}}(\rho)$: we have to measure on the probe the rank-one POVM corresponding to the Kraus decomposition $\{E_j\}_j$, and to choose the conditional correcting channels $\mathcal{C}_j^\rho$ to be equal to the unitary channels $\mathcal{C}_j^\rho(\sigma) = (U_j^\rho)^\dagger \sigma U_j^\rho$, where U_j^ρ is the unitary part of the polar decomposition $\rho E_j = U_j^\rho |\rho E_j|$. If $F_{e,\mathrm{a}}(\rho)$ is close to one, it means that the corrected channel $\sum_k \mathcal{C}_k^\rho(E_k \rho E_k^\dagger)$ acts much like the ideal channel on the support of ρ. The tricky point now is that $F_{e,\mathrm{a}}(\rho)$ *does depend* on the particular Kraus decomposition $\{E_j\}_j$, that is, on the measurement $\mathbf{M} = \{\phi_j^E\}_j$ performed upon the probe system. The natural question is which measurement $\mathbf{M}$ maximizes $F_{e,\mathrm{a}}^{\mathbf{M}}(\rho)$. We will answer showing that $F_{e,\mathrm{a}}^{\mathbf{M}}(\rho)$ essentially determines how well the measurement $\mathbf{M}$ erases from the probe the information about the input state ρ, and vice-versa. In other words, $F_{e,\mathrm{a}}^{\mathbf{M}}(\rho)$ and the erasure efficiency are equivalent measures, in the sense that the less information about the input state the measurement $\mathbf{M}$ collects, the higher the corresponding $F_{e,\mathrm{a}}^{\mathbf{M}}(\rho)$ is, and vice-versa.

Main result.— Let us now fix the input state ρ with ensemble realization $\{\rho_i\}_i$ and the probe POVM $\mathbf{M} = \{\phi_j^E\}_j$. Then $K_j = \rho^{1/2}\widetilde{\mathcal{E}}'(\phi_j^E)\rho^{1/2}/p(j)$ turn out to be normalized states, for all j, with $\sum_j p(j) K_j = \rho$. Moreover, by noticing that $\widetilde{\mathcal{E}}'(\phi_j^E) = E_j^\dagger E_j$ we can rewrite the upper bound in Eq. (4) as $F_{e,\mathrm{a}}^{\mathbf{M}}(\rho) = \sum_j p(j)\mathscr{F}(\rho, K_j)^2$, where $\mathscr{F}(\rho,\sigma) := \mathrm{Tr}[\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}]$ is the Uhlmann fidelity between two mixed states [19]. By exploiting the well-known relation [16] between the fidelity and the trace norm of the difference, defined as $\|\rho - \sigma\|_1 = \mathrm{Tr}\,|\rho - \sigma|$, that is, $\mathscr{F}(\rho,\sigma)^2 \le 1 - 2^{-2}\|\rho - \sigma\|_1^2$, together with Eq. (3), we obtain the following chain of inequalities

$$\begin{aligned} F_{e,\mathrm{a}}^{\mathbf{M}}(\rho) &\le 1 - 2^{-2}\sum_j p(j)\|\rho - K_j\|_1^2 \\ &\le 1 - 2^{-2}\sum_j p(j)\left(\sum_i |p(i) - p(i|j)|\right)^2 \\ &\le 1 - 2^{-2}\beta\sum_j p(j) D(\vec{p}(i|j)\|\vec{p}(i)) \\ &= 1 - 2^{-2}\beta I_{S:E}^{\mathbf{M}} \le 1, \end{aligned} \tag{5}$$

where $\beta = \min_i p(i) > 0$. In the second inequality we used the fact that the trace distance between two states is never smaller than the trace distance between the probability distributions obtained by measuring the same POVM $\{\rho^{-1/2}\rho_i\rho^{-1/2}\}_i$ on both states; notice that $\{\rho^{-1/2}\rho_i\rho^{-1/2}\}_i$ is a well-defined POVM on the support of ρ, since $\rho^{-1/2}\rho_i\rho^{-1/2} \ge 0$ and $\sum_i \rho^{-1/2}\rho_i\rho^{-1/2} = \mathbb{1}|_{\mathsf{Supp}(\rho)}$, [20]. In the last equality we used the trivial identity $\sum_j p(j) D(\vec{p}(i|j)\|\vec{p}(i)) = D(\vec{p}(i,j)\|\vec{p}(i)\vec{p}(j))$. Notice moreover that inequality (5) holds for every ensemble realization $\{\rho_i\}_i$ of ρ.

Equation (5) informs us that if $F_{e,\mathrm{a}}^{\mathbf{M}}(\rho)$ is sufficiently close to one (but it can be also *strictly less* than one) for a particular probe POVM $\mathbf{M}$, then the corresponding information transmission from the system to the probe is close to zero for every ensemble realization $\{\rho_i\}_i$ of ρ, that is, the measurement $\mathbf{M} = \{\phi_j^E\}_j$ is *erasing* the information about the input state ρ registered into the probe during the interaction. Equivalently, non-null information extraction always causes disturbance on the input ensemble, even allowing input-dependent environment-assisted correction schemes. Our approach hence embodies a quite general *information-disturbance tradeoff*. It is worth stressing here that even if we drop the complete controllability hypothesis, the same conclusions are true, because the information extracted by the "uncontrolled" POVM $\{\phi_j^P \otimes \mathbb{1}_R\}_j$ is clearly less than or equal to the information extracted by a fully controlled POVM $\{\phi_j^E\}_j$: averaging decreases information.

The converse argument runs as follows. We have to check that, given the channel and its realization as an interaction with a probe, for a suitable probe rank-one POVM $\mathbf{M} = \{\phi_j^E\}_j$, the information transmission is poor *for all possible* ensemble realizations $\{\rho_i\}_i$ of a given input ρ. Luckily enough, we can restrict our attention to just one particular set $\{\rho_i\}_i$ of input states, with $\sum_i \rho_i = \rho$, being also *informationally complete*, that is, such that every operator O on the support of ρ is uniquely defined by its expectation values on such ensemble. The existence of this kind of ensembles for every finite dimension has been constructively proved in Ref. [21]. The reconstruction formula then holds

$$O = \sum_i \mathrm{Tr}[O\ \rho^{-1/2}\rho_i\rho^{-1/2}]\rho_i', \tag{6}$$

where the operators $\{\rho_i'\}_i$ are limited and hermitian but neither positive definite nor semi-definite, in general. By exploiting the well-known inequality $\mathscr{F}(\rho,\sigma)^2 \geq 1 - \|\rho - \sigma\|_1$, [16], we have the following relations

$$\begin{aligned} F_{e,\mathrm{a}}^{\mathbf{M}}(\rho) &\geq 1 - \sum_j p(j)\|\rho - K_j\|_1 \\ &= 1 - \sum_{ij} p(j)\|p(i)\rho_i' - p(i|j)\rho_i'\|_1 \\ &\geq 1 - |\Gamma| \sum_{ij} p(j)|p(i) - p(i|j)| \\ &\geq 1 - \sqrt{2}|\Gamma|\sqrt{I_{S:E}^{\mathbf{M}}}, \end{aligned} \tag{7}$$

where $|\Gamma| = \max_i \|\rho_i'\|_1 < \infty$. In the first equality we used the reconstruction formula (6). Notice that, if $I_{S:E}^{\mathbf{M}}$ is sufficiently close to zero for one particular informationally complete input ensemble realization $\{\rho_i\}_i$, then $F_{e,\mathrm{a}}^{\mathbf{M}}(\rho)$ is close to one, and, by Eq. (5), $I_{S:E}^{\mathbf{M}}$ is also close to zero for all possible input ensemble realizations of ρ. The implications then turn out to be equivalences [22].

We can hence conclude, by stating that *for every channel realized as an interaction of an input state ρ with a probe, even if perfect quantum erasure is impossible, the more a given POVM erases from the probe the classical information which can be carried by the input state and get stored in the probe during the interaction, the closer (on the support of ρ) the corresponding corrected channel is with respect to the ideal one, and vice-versa.* To find the optimal erasure measurement for a given channel and a given input state remains an open problem. Incidentally, it is worth noticing that if we consider $\rho = \mathbb{1}/d$—indeed just an invertible ρ suffices—, then the quantum capacity of the corresponding corrected channel is maximum over the whole input Hilbert space $\mathscr{H}_S$.It is possible to achieve *perfect* erasure, that is $F_{e,\mathrm{a}}^{\mathbf{M}}(\mathbb{1}/d) = 1$, if and only if the channel admits a random-unitary decomposition [2], that is, $\mathcal{E}(\rho) = \sum_j p(j)U_j\rho U_j^\dagger$, for some probability distribution $\vec{p}(j)$ and unitary operators $\{U_j\}_j$. In this case, for the corresponding measurement $\mathbf{M}$, $I_{S:E}^{\mathbf{M}}$ is rigorously zero *for every* possible input ensemble [23].

Acknowledgments.— Discussions with M Hayashi, L Maccone, and M F Sacchi are gratefully acknowledged. The author also thanks the anonymous Referees whose comments greatly contributed in clarifying the manuscript. Japan Science and Technology Agency is acknowledged for support through the ERATO-SORST Quantum Computation and Information Project.

[*] Electronic address: buscemi@qci.jst.go.jp

[1] K Horodecki, M Horodecki, P Horodecki, J Oppenheim, Phys. Rev. Lett. **94**, 160502 (2005); K Horodecki, M Horodecki, P Horodecki, D Leung, and J Oppenheim, preprint arXiv:quant-ph/0702077v1.

[2] M Gregoratti and R F Werner, J. Mod. Opt. **50**, 915 (2003).

[3] M O Scully and K Drühl, Phys. Rev. A **25**, 2208 (1982); S Dürr, T Nonn, and G Rempe, Nature **395**, 33 (1998).

[4] B-G Englert, Phys. Rev. Lett. **77**, 2154 (1996).

[5] B-G Englert and J Bergou, Opt. Commun. **179**, 337 (2000).

[6] C H Bennett, G Brassard, C Crépeau, R Jozsa, A Peres, and W K Wootters, Phys. Rev. Lett. **70**, 1895 (1993).

[7] S Popescu, Phys. Rev. Lett. **72**, 797 (1994); N Gisin, Phys. Lett. A **210**, 157 (1996); L Mišta, Jr. and R Filip, Phys. Rev. A **71**, 022319 (2005).

[8] K Kraus, *States, Effects, and Operations: Fundamental Notions in Quantum Theory*, Lect. Notes Phys. **190** (Springer-Verlag, Berlin, 1983).

[9] W F Stinespring, Proc. Am. Math. Soc. **6**, 211 (1955).

[10] M Ozawa, J. Math. Phys. **25**, 79 (1984).

[11] B Schumacher and M D Westmoreland, Quant. Inf. Processing **1**, 5 (2002).

[12] A S Holevo, Prob. Th. Appl. **51**, 133 (2006).

[13] F Buscemi, G M D'Ariano, M Keyl, P Perinotti, and R F Werner, J. Math. Phys. **46**, 082109 (2005).

[14] E B Davies and J T Lewis, Commum. Math. Phys. **17**, 239 (1970).

[15] T M Thomas and J A Cover, *Elements of Information Theory* (John Wiley and Sons, New York, 1991).

[16] M Hayashi, *Quantum Information: an Introduction* (Springer-Verlag, Berlin, 2006).

[17] K M R Audenaert and J Eisert, J. Math. Phys. **46**, 102104 (2005).

[18] B Schumacher, Phys. Rev. A **54**, 2614 (1996).

[19] A Uhlmann, Rep. Math. Phys. **9**, 273 (1976).

[20] L P Hughston, R Jozsa, and W K Wootters, Phys. Lett. A **183**, 14 (1993).

[21] G M D'Ariano, P Perinotti, and M F Sacchi, J. Opt. B: Quantum and Semicl. Optics **6**, S487 (2004).

[22] Formally speaking, the two quantities $(1 - F_{e,\mathrm{a}}^{\mathbf{M}}(\rho))$ and $\mathscr{I}(\rho,\mathbf{M}) := \max_{\{\rho_i\}_i:\sum_i \rho_i = \rho} I_{S:E}^{\mathbf{M}}$ are equivalent, since $a(\rho)\mathscr{I}(\rho,\mathbf{M})^2 \leq (1 - F_{e,\mathrm{a}}^{\mathbf{M}}(\rho))^2 \leq b(\rho)\mathscr{I}(\rho,\mathbf{M})$, with $0 < a(\rho) < b(\rho) < \infty$.

[23] F Buscemi, Phys. Lett.A **360**, 256 (2006).